

The Grassmannian Atlas: A General Framework for Exploring Linear Projections of High-Dimensional Data

EuroVis 2016, U Utah

背景与动机

- 对象：线性投影（linear projection）
- 问题：目前很多方法（grand tour, projection pursuit）找出的一组投影中，很多投影所表达的信息是相似的，比如两个投影角度之间的角度差很低，或是两个投影结果实际上是旋转相似的。
- 目标：找一组好的投影角度，以最大化角度之间的信息差，即用最少的投影结果获得最全面的信息。

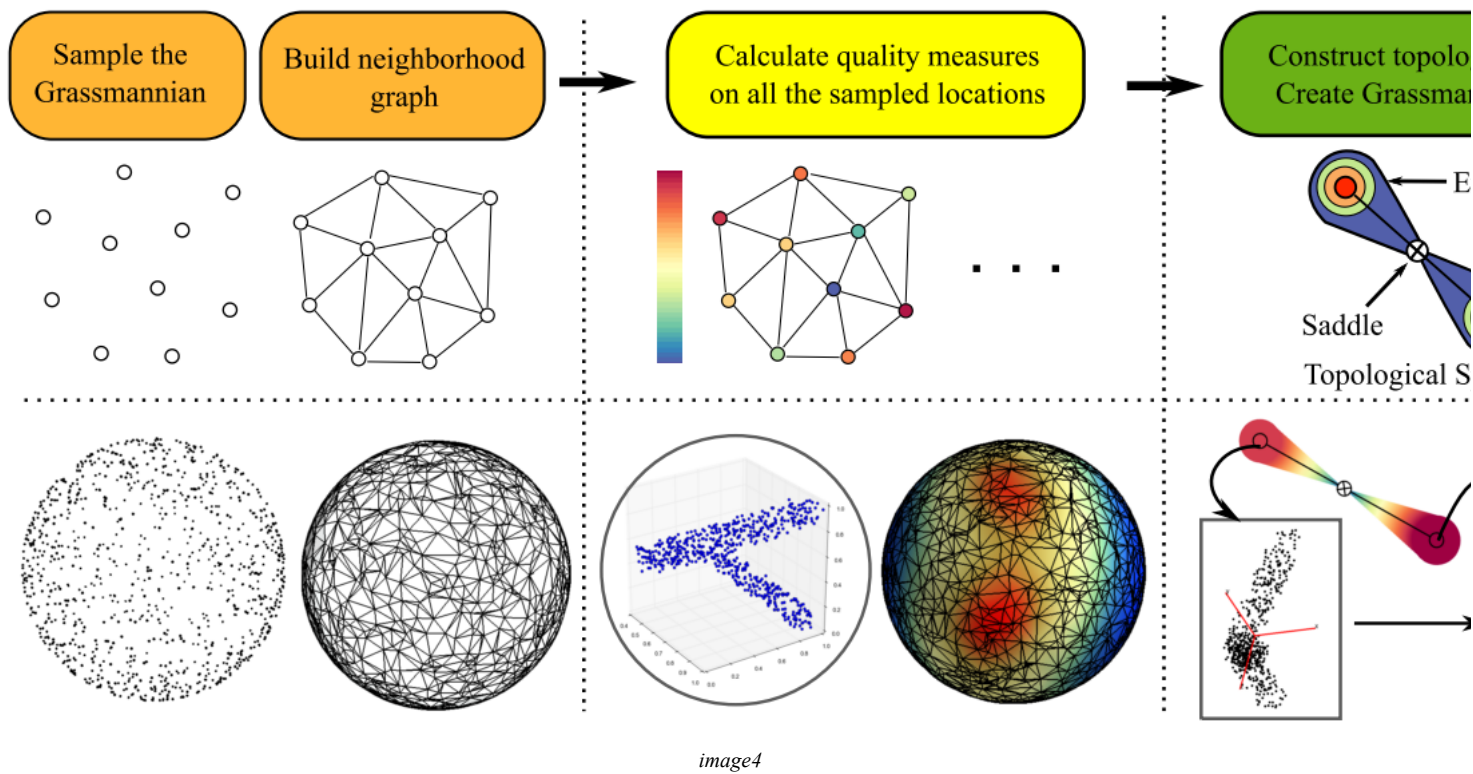
相关工作

- InfoVis 2015 Optimal sets of projections of high-dimensional data，目标与本文类似。

贡献

- 使用格拉斯曼流型描述所有的linear projection，并在其上选择合适的投影角度。
- Topological spine，一种表达投影角度质量分布的视觉设计。
- 交互式系统

概述



- 工具：格拉斯曼流型（Grassmannian）
 - 流水线
- 数据集外表面采样出一层壳（格拉斯曼流型）
 - 建立这些采样点的k-NN graph。
 - 计算所有采样点角度上的投影质量指标。
 - 构建topological spine，用于展示整个k-NN graph上的投影质量分布情况。

方法

格拉斯曼流型采样

数据集上采出Grassmannian $Gr(2, n)$ ，并构建k-NN graph。

投影质量指标

主要用三类： - scagnostics - projection pursuit index - MDS、LDA、PCA等方法中使用的目标评价函数

投影质量分布的拓扑简化

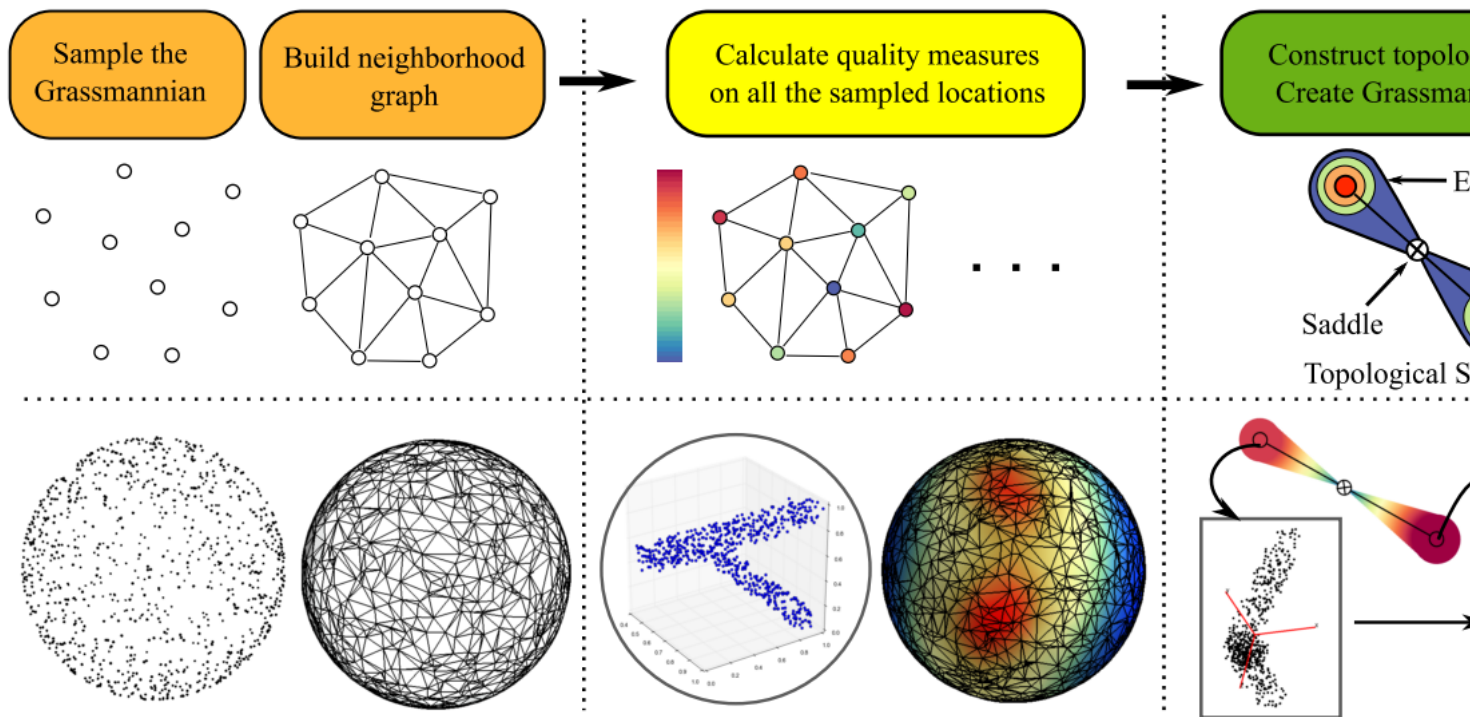
- 构建Morse complex，用于描述一个函数的拓扑结构摘要。
- 使用topological spine可视化整个linear projection空间。
- topological spine来自TVCG 2011： Topological spines: A structure-preserving visual representation of scalar field

交互界面

在spine上选择local maxima，投影视图中展示对应的数据点线性投影结果。

实验

使用多种投影指标，可以快速找到符合指标的投影结果，辅助分析者发现一些已有方法难以快速找到的特征，并大幅减少所需查看的投影数量。例如下图对outlier的探索。



感想

- 投影都是流形学习，数学工具很重要。
- 线性投影找好的角度这一问题很老但一直没有很好地解决。近两年有几篇相关文章出现，应该是有人在深挖。
- 可能可以结合subspace探索，碰撞出新想法。